SEGMENT-BASED STEREO-MATCHING VIA PLANE AND ANGLE SWEEPING

Cevahir Çığla[1], Xenophon Zabulis[2] and A. Aydın Alatan[1]

[1]Department of Electrical and Electronics Engineering, M.E.T.U, Turkey
[2]Informatics and Telematics Institute, Thessaloniki, Greece
e-mail: cevahir@eee.metu.edu.tr, xenophon@iti.gr, alatan@eee.metu.edu.tr

ABSTRACT

A novel approach for segment-based stereo matching problem is presented, based on a modified plane-sweeping strategy. The space is initially divided into planes that are located at different depth levels via plane sweeping by the help of region-wise planarity assumption for the scene. Over-segmented homogenous color regions are utilized for defining planar segment boundaries and plane equations are determined by angle sweeping at different planes. The robustness of depth map estimates is improved by warping segments into the other image via the resulting homographies. In order to refine the reconstruction quality and update segment depths, as well as plane normals, with smoothness and visibility constraints, a greedy search algorithm is applied. Based on the simulation results, the proposed algorithm handles large un-textured regions, depth discontinuities at object boundaries and slanted surfaces. Moreover, the algorithm could be easily upgraded from stereo to multi-view case, since 3D plane equations are already determined.

Index Terms**—** segment based stereo matching, plane sweeping, angle sweeping.

1. INTRODUCTION

Dense depth-map estimation has attracted many researches due to its wide application areas in computer vision, such as 3D object modeling, segmentation and image-based rendering. In literature, there are many algorithms on extraction of depth information from stereo images [1]. However, the problem still remains challenging due to the variety in the scene complexity and the limitation of the observed data. Most of the dense depth estimation algorithms try to solve the problem based on two basic assumptions, namely the *smoothness of the depth field* and the *high level of visual similarity* between image neighborhoods of corresponding pixels. Considering this fact, dense matching algorithms can be generally classified into two groups, as *local* and *global* techniques [1]. Local methods are based on searching the depth-space resulting in the best intensity match by utilizing "winner take all" optimization. On the other hand, global methods consider the smoothness constraint and the depth-map is extracted by minimizing a cost function via global optimization techniques, such as *graph cuts* [2] or *dynamic programming* [3].

Dense depth map of a scene should smoothly vary on the surfaces of the objects and change sharply at the object boundaries. Segment-based stereo matching algorithms [4,5,6,7,8] developed in the last decade, are capable of handling such smoothness variations at object boundaries, which result in realistic depth maps. These algorithms rely on the assumption that the scene is composed of small non-overlapping planes, all of which correspond to distinct segments obtained via grouping pixels of homogenous color. Hence, smoothness constraint is valid within each segment and depth distribution is allowed to change sharply between segment boundaries, which generally correspond to object boundaries as well. Segment-based stereo matching algorithms mainly consists of four steps: In the first step, the reference image is over-segmented, in order to obtain non-overlapping plane masks. After the segmentation stage, the initial depth map is estimated by local stereo matching algorithms for both of the images and the reliable points are determined. In the third step, plane parameters are estimated for each segment by the utilization of reliable points. In the final step, the depth (disparity) distribution among segments is achieved by different optimization techniques [4,5,6,7,8].

This paper proposes a new approach to segment-based stereo matching algorithms, based on a modified plane-sweep method [9]. In addition, segments are treated as feature points and the best plane representation is determined for each segment, instead of estimating depth values for each pixel. With this new approach, 3D plane normal estimation by angle-sweeping [10] replaces the plane parameter estimation step. The details of the proposed algorithm are explained in Section 2, whereas the results are presented in Section 3. The last section is devoted to the conclusion and future directions.

2. ALGORITHM

The proposed algorithm is composed of three steps, segmentation of the stereo images, initial depth-plane estimation and iterative update of the plane parameters. In the following subsections, each of these steps is explained in detail. The algorithm takes two calibrated images as input and the 3D location of each pixel is determined at the final stage by the help of the estimated plane equations of segments.

3.1 Segmentation

The main assumption of the algorithm is modeling of the scene with planar patches. These patches should correspond to sub-regions of homogenous color, named *segments*. In addition, depth variation is assumed to be smooth within each segment, as well as between neighboring segments, which have similar color distributions. Therefore, depth discontinuities should only be observed only at segment boundaries which have high intensity differences. Moreover, segmentation should unite similar colored pixels to handle texture-less regions. In this work due to its logical segmentation of regions having low intensity gradient, the *mean-shift image segmentation* algorithm [11] is utilized. The results of such segmentation on "*Teddy*" image sequence are illustrated in Fig.1.

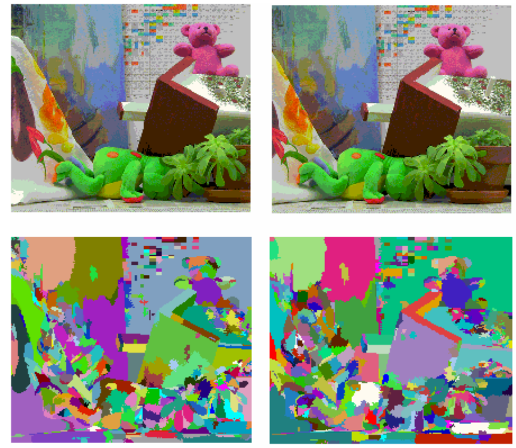

Figure 1 : Stereo images (left and right views) are segmented by mean-shift segmentation algorithm, respectively.

3.2 Initial Depth Estimation

In all the segment-based stereo algorithms, initial depth estimation step is performed via local matching in the pixel domain which is followed by estimation of the planar models within segments. In our method, a new initial depth estimation approach is proposed by treating all segments, as feature points. Instead of matching pixels individually, group of pixels is matched to a region, independent to the shape of the region, since it increases the robustness of the algorithm against noise and repeated regions. There are three steps in the proposed initial depth estimation method. In the plane sweeping step, 3-D space is limited for a detailed search, compared to the angle sweeping step. After the sweeping operations on both of the images, initial depth is determined in the third step.

3.2.1. Plane Sweeping

Any 3-D space can be divided into parallel planes [9] which are located at different depths, yielding to different disparities between the stereo image pair. The separation of the planes is performed along the principal axis of the reference camera, which is also assumed to be the z-axis (see Fig. 2) and plane equations are determined based on the camera calibration parameters and disparities. Once the planes are determined, the relation between stereo images among a plane can be defined as a homography which is explicitly given in (1) and (2)

$$H_i = [P_1^i\ P_2^i\ P_4^i] - P_3^i\,.[n_1/n_3\ n_2/n_3\ n_4/n_3] \qquad (1)$$

$$H_{12} = H_1^{-1}.H_2 \qquad (2)$$

where P_j^i corresponds to the j^{th} column of the projection matrix of the i^{th} camera and n_i indicates the plane normal.

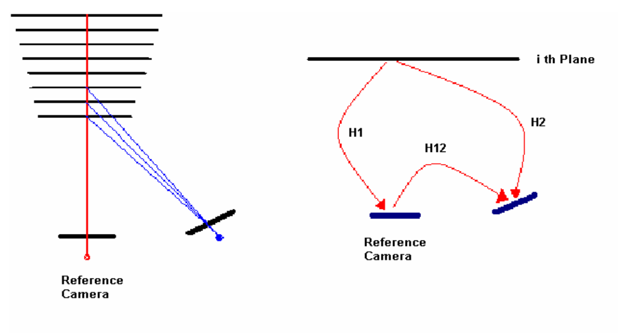

Figure 2 : The space is divided into non-intersecting planes. Image planes are related via homography via such planes.

For each segment, a search is performed on depth space by the utilization of the corresponding homographies. The cost function of ith segment (S_i) on the jth plane is defined as follows:

$$C_{ij} = \frac{1}{N}\sum_{(m,n)\in S_i} \left| I_1(m,n) - I_2(m',n') \right| \qquad (3)$$

where $(m',n',1) = H_{12}.(m,n,1)$ and N is the number of pixels in S_i. The planes are sorted according to the resulting cost value for each segment independently and the first n best planes are utilized during the angle sweep step.

3.2.2. Angle Sweeping

In the previous step, the planes are constructed in the direction of the principle axis. This property is valid for small segments; however as the segment size increases the assumption will fail, if there are slanted surfaces in the scene, which is a general case. In order to handle such situations, large segments are rotated in *x*- and *y*-directions according to their centeroids (see Fig. 3). The rotation is performed within the angle range of the viewing rays from the camera centers to the centeroids of the segment in 3D according to the corresponding depth since the segment should be observed from both of the images properly.

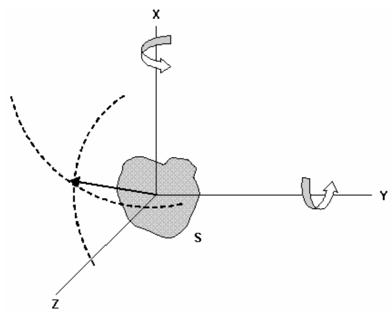

Figure 3 : Segments are rotated from centeroids in X and Y directions

Rotation of a plane results in a different homography between stereo images, which can still be obtained via (1) and (2), after accordingly modulating the plane normal. The search in depth space is performed by seeking for the best rotation and depth combination after sweeping the normal of the segments at different angles. The combination that gives the minimum cost is chosen as the location of the segment in 3D. The rotation of a segment does not violate the smoothness between the pixels belonging to the segment, since the depth should change according to plane equation.

3.2.3. Finalizing Initial Depth Map

So far, initial depth estimates for the segments in both views have been extracted. These estimates are utilized to perform a left-right consistency check and determine the reliable segments in the reference view. The reliable segments are determined via consistency of the pixels within the segment. After the pixel-wise cross-check in the reference view, if the number of consistent pixels is greater than a threshold, **α**%, (**α**=70 in our experiments) of the size of the segment, then the segment is assumed to be consistent. Once the consistent depth map (for reference view only) is obtained, another search for the inconsistent segments is performed.

In this approach, it is assumed that only one segment has the pixel matched to a specific pixel in the other image. The second view is reconstructed with the consistent segments and the search for inconsistent segments is performed among free regions of the reconstructed view by applying same sweeping operations. At this step, visibility of pixels is also considered; the cost of a segment (inconsistent) is calculated among the visible pixels on the reconstructed view.

The resulting depth maps of "Teddy" images estimated at each step are illustrated in Fig 4; (a) and (b) are the depth maps of both images that is determined by plane sweeping only; (c) and (d) are the depth maps after angle sweeping at large segments, (e) is the depth field of consistent segments, in which the red colored pixels indicate the inconsistent regions, and (f) is the resulting depth map, which is obtained before the iterative final stage. Since the resultant depth map is obtained by only intensity matching, there can be some regions which have discontinuous depth distribution.

3.3 Iterative Plane Update

In this step, the plane equations of the segments are updated by considering the smoothness and visibility constraints in order to refine the depth map. The reliability of the depth map is measured with the reconstruction quality of the other image pair from the reference image.

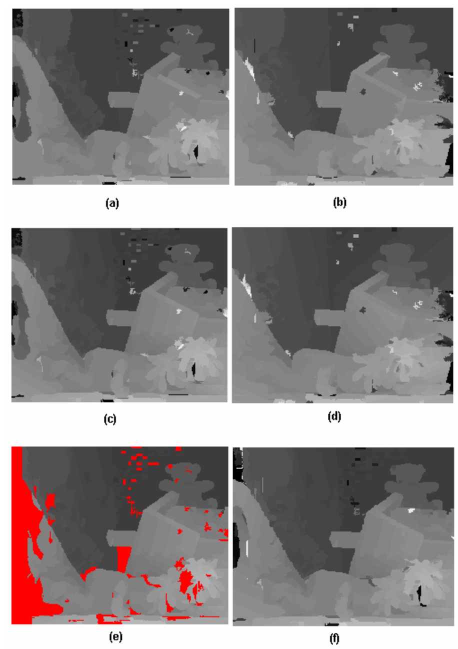

Figure 4 : The resultant depth maps after each step of the initial depth estimation.

In order to handle the assumptions above, the cost function is updated with a smoothness term that takes depth differences of the neighboring pixels on the segment boundaries into account. One of the images is warped onto the other pair by utilizing plane equations and the resulting homographies between two images. During such a reconstruction, some pixels have more than one correspondence from the pixels on the reference image. In this case, the pixel, which is closer to the camera, is rendered. For realizing the visibility, the reconstructed image is stored in a Z-buffer and the pixels on the top of Z-buffer are utilized for the intensity filling [5]. The intensity similarity is measured among the visible pixels; hence the new cost function is obtained as:

$$C = \frac{1}{N}\sum_{X \in V_i} \left| I_r(x) - I_2(x) \right| + \lambda \sum_{\substack{X \in B_i \\ X' \in N_i}} \left| D(x) - D(x') \right| \qquad (4)$$

where D is the depth map, I_r is the reconstructed image, V_i is the set of visible of i^{th} segment, B_i is the boundary and N_i is the neighboring pixels of the segment.

The solution of the above optimization problem is np-complete and there are different types of solutions, such as

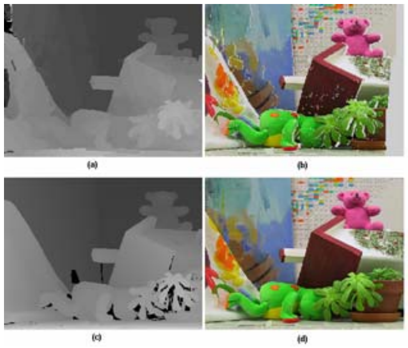

Figure 5 : (a) resultant depth map, (b) reconstructed right image, (c) the ground truth depth map, (d) the original right image

graph-cut [8], belief propagation [6]. In our algorithm, we utilize a method, which is similar to the greedy search algorithm given in [12]. In this method, for each segment, a search is performed in the depth space bounded with the depth planes of its neighboring segments. If the segment is considerably large, angle sweeping is also applied within the bounded region. The model which gives the best improvement in the cost function is assigned to the segment; however the models are updated after all of the segments are visited. This operation is performed iteratively until there is sufficiently small number of updates among all segments.

4. RESULTS

We tested our algorithm on the stereo image data set given in [13]. In addition, *Microsoft Break-dancer* multiple image sequence [14] has also been utilized for observing the performance of the algorithm. In Fig. 5, the estimated depth map of the *Teddy* image and the reconstructed right image are illustrated. In the resultant depth map, depth discontinuities on object boundaries can be observed clearly and the depth of the untextured regions are also estimated.. The depth maps, as well as the reference images for different time instants of *Break-dancer* are shown in Fig. 6.

5. CONCLUSION AND FUTURE WORKS

A new approach to segment-based depth estimation algorithms has been proposed via plane and angle sweeping in 3D. In addition, arbitrary-shaped segment matching is performed, instead of pixel matching, which reverses the pixel-to-segment methods. The untextured regions and discontinuities at object boundaries are handled precisely. In addition, the algorithm can easily be adapted to the multi-view case, which is a future work of our study. The algorithm can also be extended for higher order models, such as circular or parabolic surfaces which will improve the depth map quality especially for the scenes where the planar assumption does not hold.

Figure 6: (a),(c) estimated depth maps of the reference images; (b),(d) reference images at different time instants

6. ACKNOWLEDGEMENT

This work is funded by EC IST 6th Framework 3DTV NoE and partially funded by TÜBİTAK under Career Project 104E022.

7. REFERENCES

[1] D. Sharstein and R. Szeliski, *A taxonomy and evaluation of dense two-frame stereo correspondence algorithms*. In International Journal of Computer Vision, volume 47, pages 7-42, April 2002

[2]V.Kolmogorov and R.Zabih, *Computing visual correspondence with occlusions using graph cuts*. In ICCV, volume II, pages 508-515, 2001.

[3] G. Van Meerbergen, M. Vergauven, M.Pollefeys, and Luc Van Gool, *A Hierarchical Symetric Stereo Algorithm using Dynamic Programming,* In International Journal of Computer Vision 47(1/2/3), 275-285, 2002

[4] C. Lei, J. Selzer, and Y. Yang. *Region-tree based stereo using dynamic programming optimization*. CVPR 2006.

[5] M. Bleyer and M. Gelautz. *A Layered stereo algorithm using image segmentation and global visibility constraints*. In ICIP 2004.

[6] A. Klaus et all. *Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure*. ICPR 2006.

[7] Sang Yoon Park, Sang Hwa Li, and Nam Ik Cho, *Segmentation Based Disparity Estimation Using Color and Depth Information,* In ICIP 2004

[8] L. Hong and G. Chen, *Segment-based stereo matching using graph cuts*, in CVPR, 1, pp. 74-81, 2004.

[9] Robert T.Collins, *A Space-Sweep Approach to True Multi-Image Matching*. In Proc. CVPR96, pages 358-363, 1996

[10]Xenophon Zabulis ans Kostas Daniilidis, *Multi-camera reconstruction based on surface normal estimation and best viewpoint selection*. In Proceedings of the 2nd International Symposium on 3DPVT, 2004.

[11] D. Comaniciu and P. Meer. *Mean shift: A robust approach toward feature space analysis*. *IEEE:PAMI*, 24(5):603–619,May 2002.

[12] H. Tao and H. Sawhney, .*Global matching criterion and color segmentation based stereo*,. in *WACV*, 2000, pp. 246.253.

[13] *http://cat.middlebury.edu/stereo/data.html*

[14] http://research.microsoft.com/IVM/3DVideoDownload/